Sigmoid Neuron

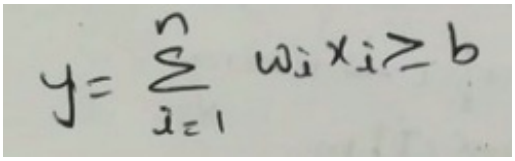
medium.com/@manveetdn/notes-on-sigmoid-neuron-padhai-onefourthlabs-course-a-first-course-on-deep-learning-8f189b8e368f

> **Disclaimer: This is notes on "Sigmoid Neuron" Lesson (PadhAI onefourthlabs course "A First Course on Deep Learning")**

Sigmoid Neuron is the Building Block and foundation of Deep Neural Network.

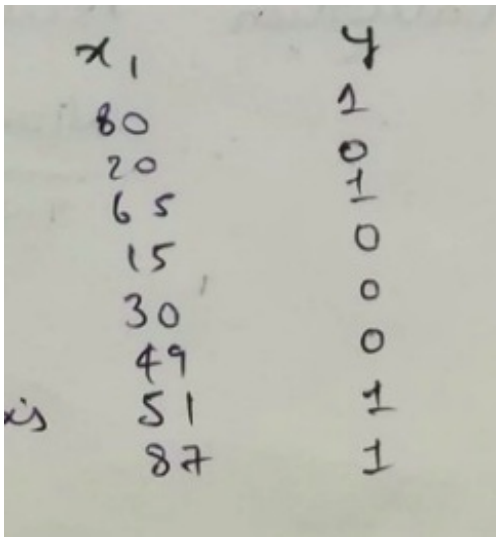
Limitations of perceptron:

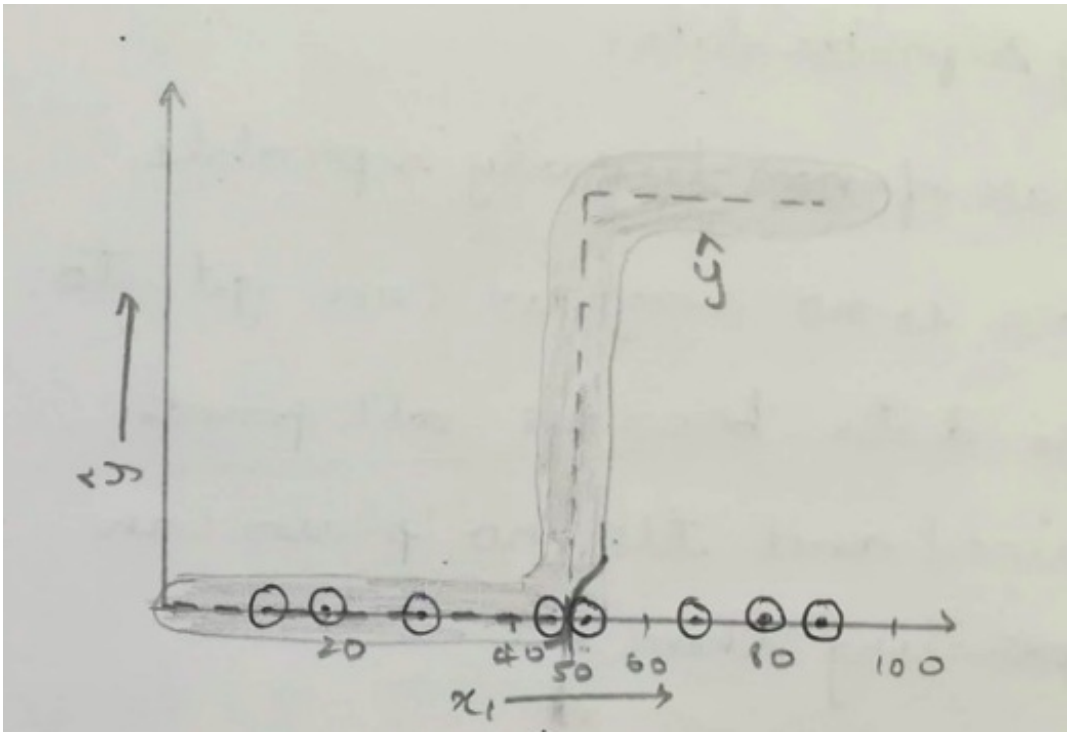
This is the perceptron equation.If are dealing with only one variable x1 and final output is y.



Suppose w1 = 0.2, b = -10

The plot we get will be on x-axis.The plot will be as follows.

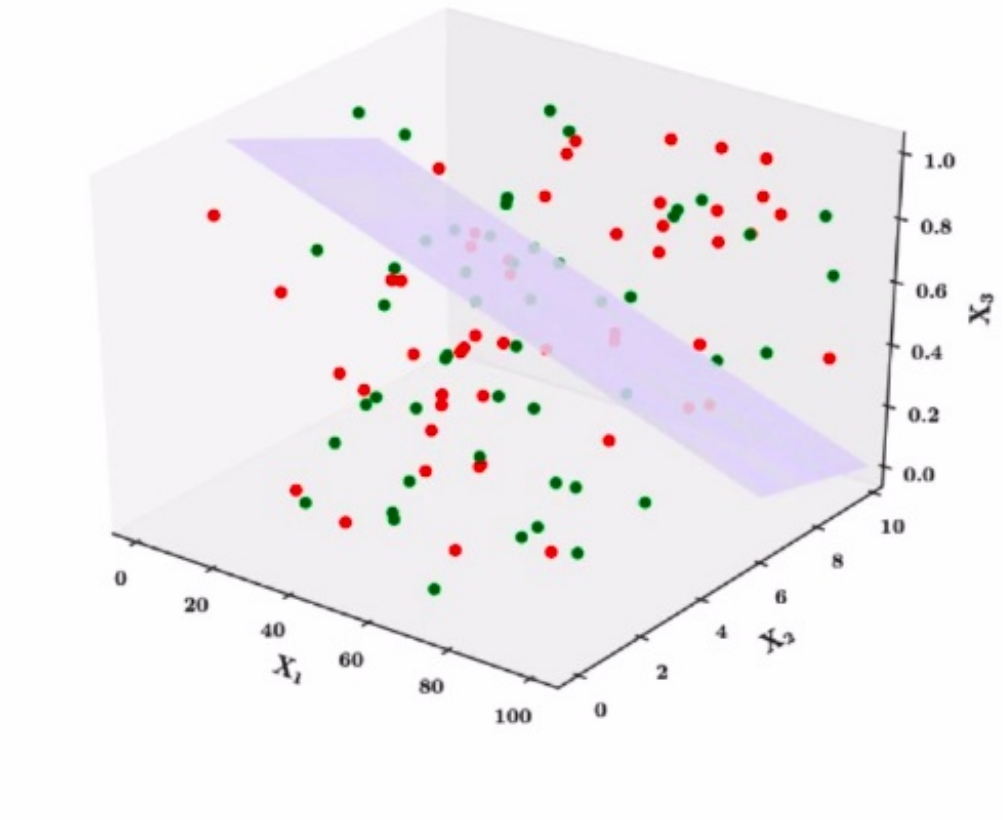




yhat will the seprating Function if we have x1 and x2 2D data here x2 intercept y value will be -(b/w2).

But here we just have x, 1D data.Therefore, its intercept = -(b/w1)

Here, the perceptron algorithm is dividing into two halves positivies and negatives that is above diagram at point 50.



Perceptron model fails when we have non linearly separable data like this

Same case with 2D data It will be seperating into 2 halves postives and negatives

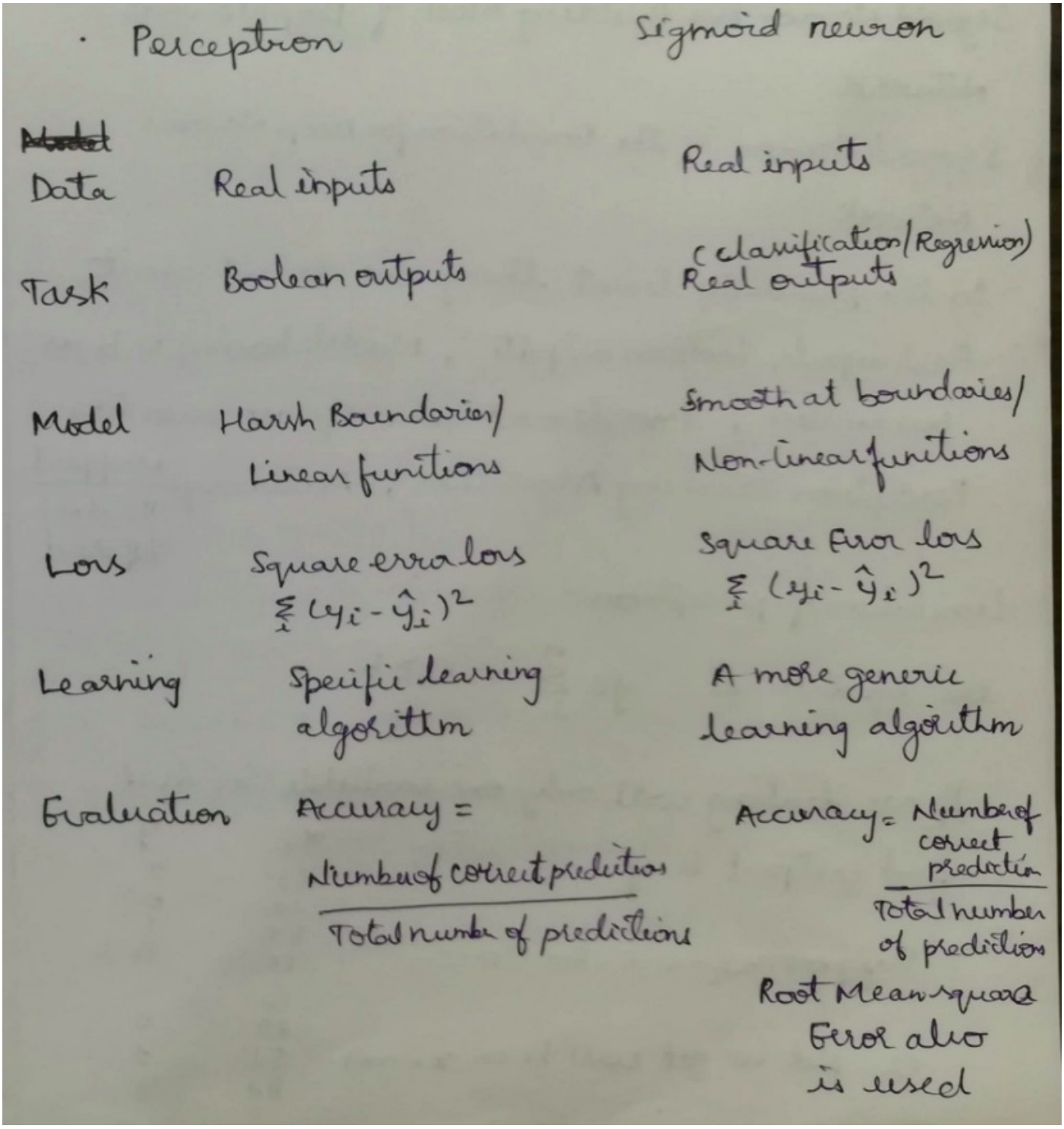
Non Linearly Separable data:

In case non linearly separable data there is no way we can get t seperate the data because all points will be mixed and the no plain can satisfy seprating them.

Therefore, Perceptron cannot deal with deal with non Linearly separable data.

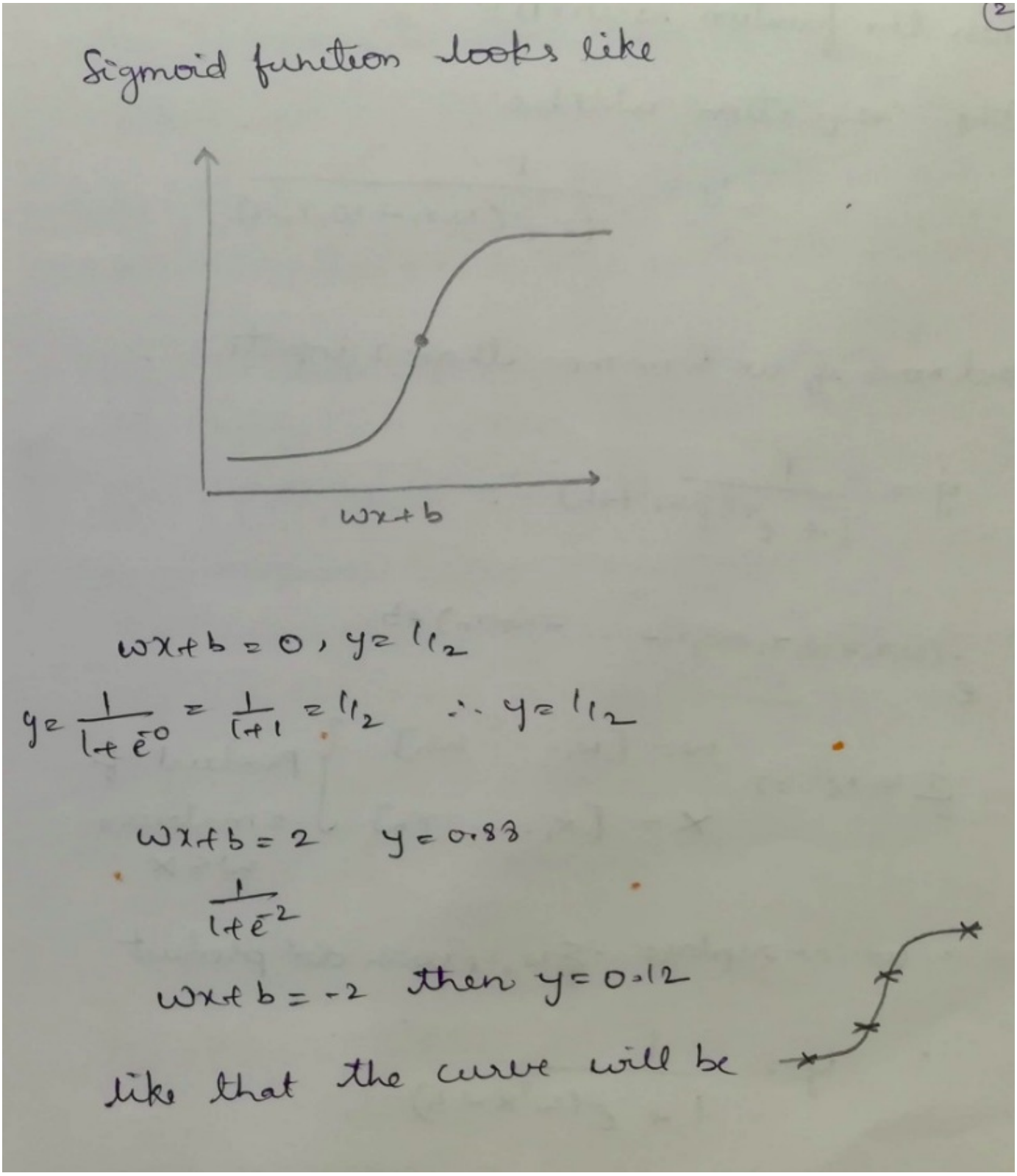
Everyone more disadvantage of perceptron is output are zero or one there is nothing intermediate value possible.

Difference between Sigmoid Neuron and Perceptron



6 jars of Sigmoid:

Model:



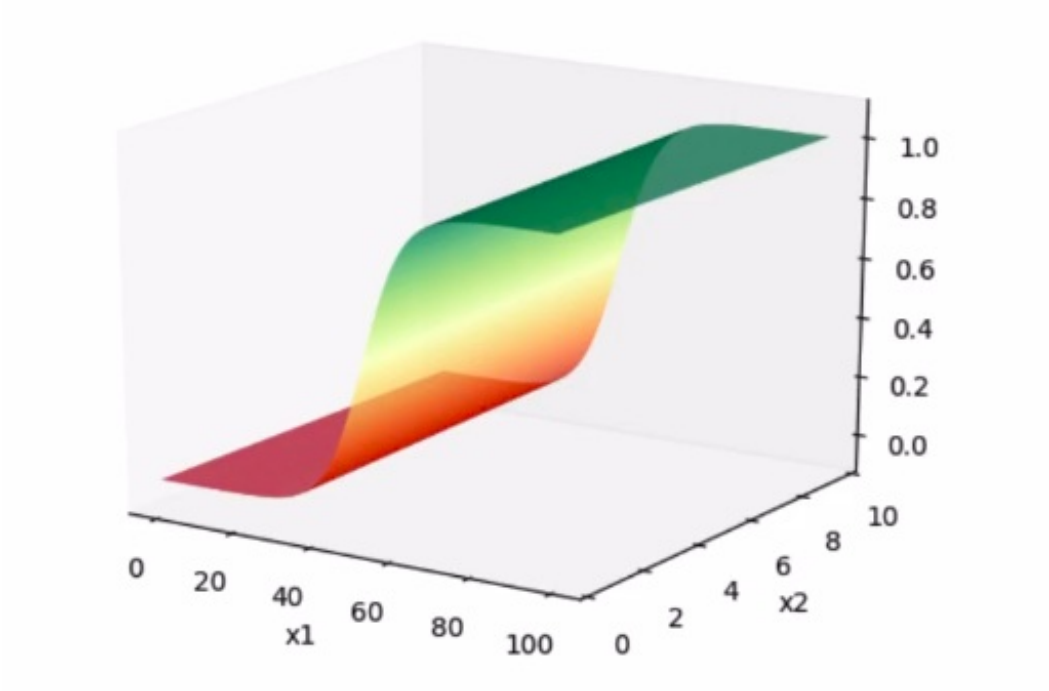
Sigmoid Function graph

The beside one is the logistic function y = 1/1+e^-(wx+b)

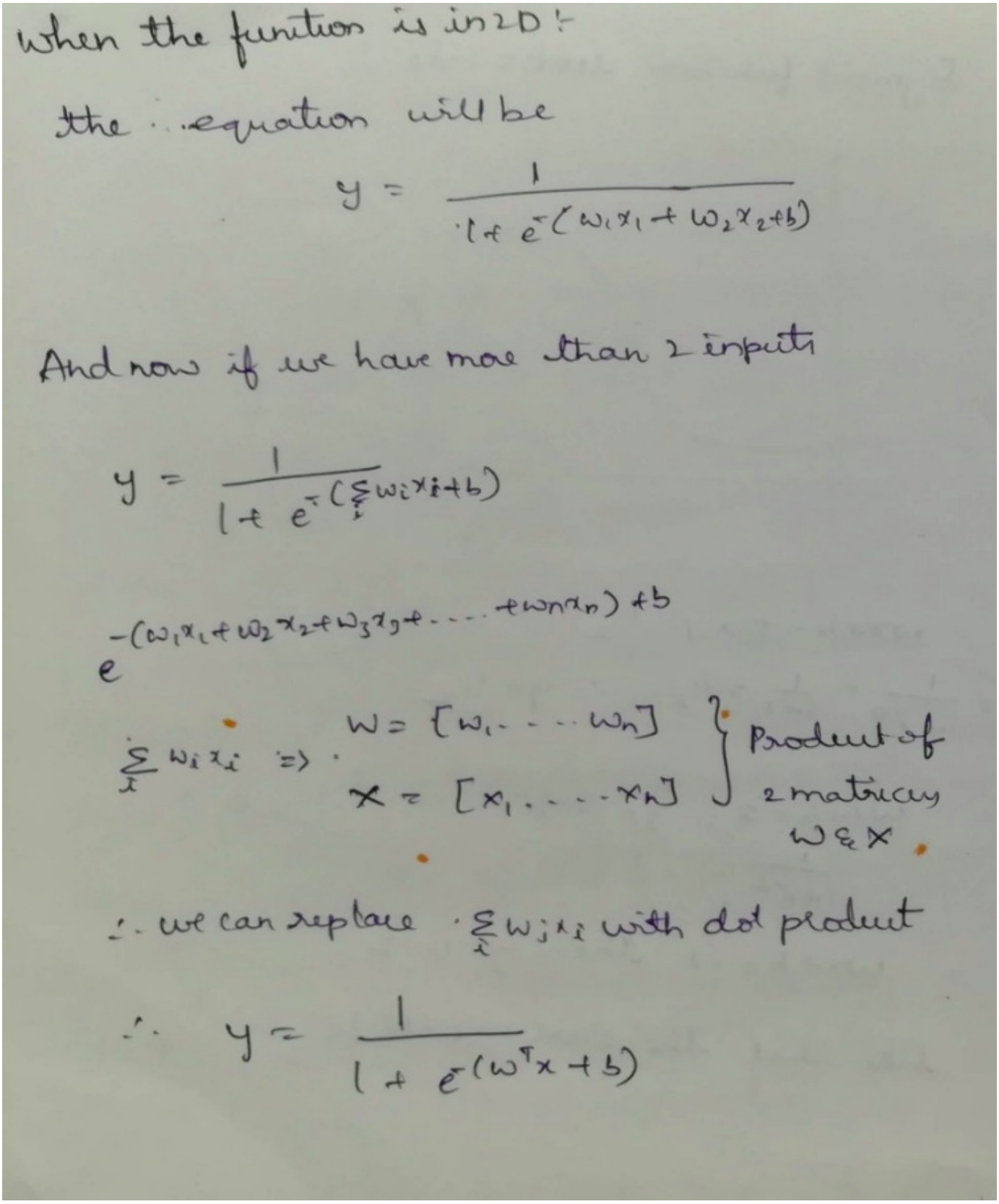
If I dont ave any from like if y =1 Some decision and if y = 0 for Some other ecision.

If we give x and y if x belongs to R we will evaluate above function.

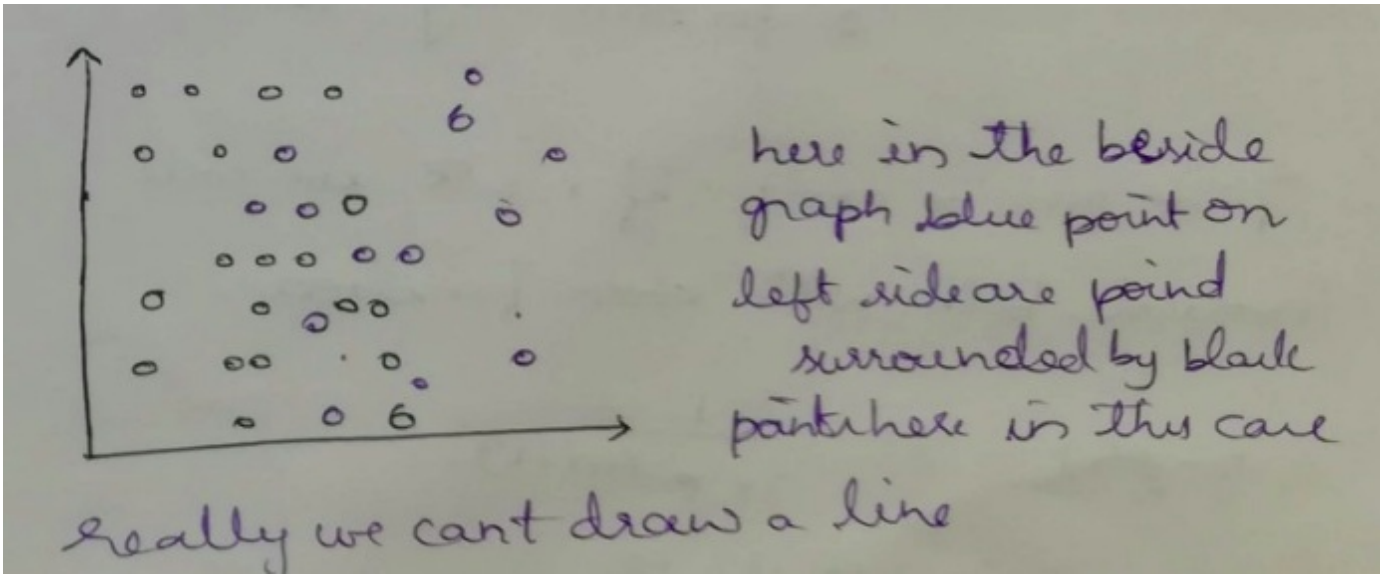
y = f(x), y = 1/1+ e^-(wx+b) wit this this function.If the value of wx+b = 0 then we will be getting the output y value 0.5. Similarly, for wx+b = 2 we get y = 0.88 and for wx+b = -2 we get y = 0.12 like that. If we plot a graph, we get the shape as the beside graph.



Sigmoid Function In 3-D space.

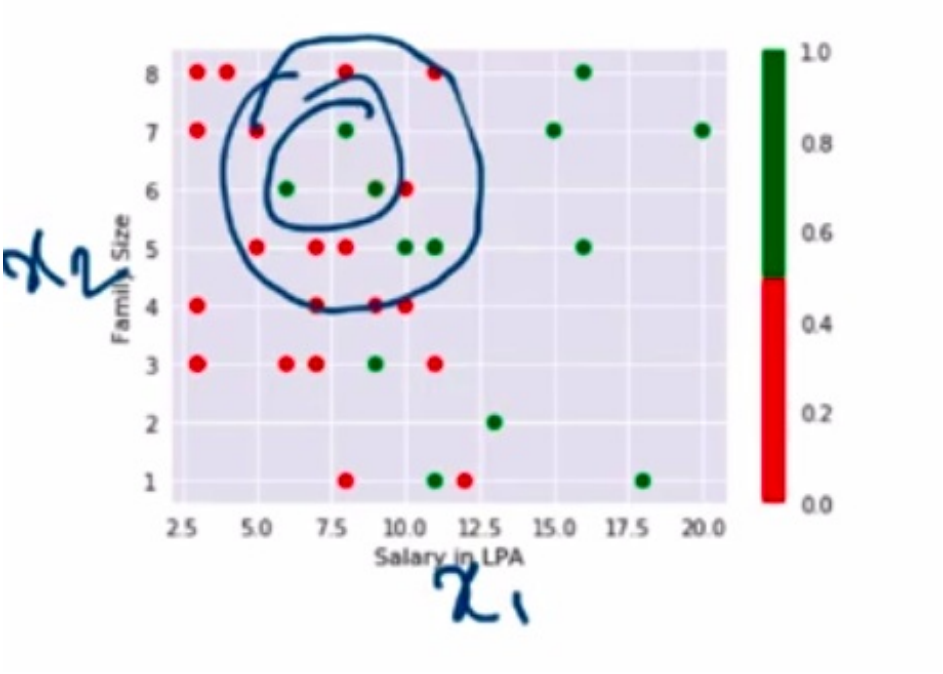


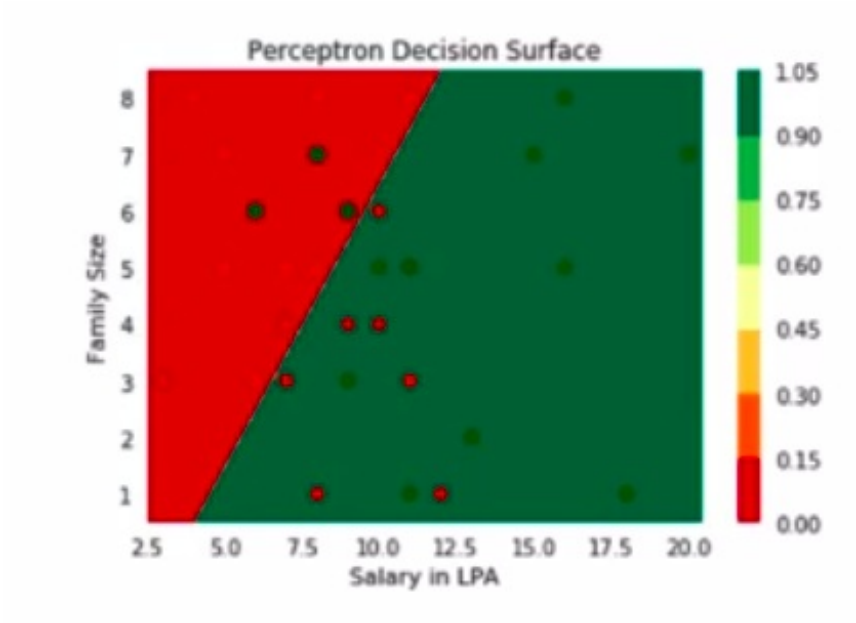
Here as shown in the left side y increase with increase in the values of x. the graph will grow in a non linear from but it will increase in a curve in a curve shape.



Here even if the graph is like this some time in case of non linear separable data, we may fail for example if we have a graph link the below with plotted points.Here we really cant darw the line.

In the case of 3D sigmoid function as shown below.





If we have data like this which is non-linearly separable as in fig(1) and the green points are surrounded by red points as in fig(2) its very difficult to classify and perceptron model classifies at max level as shown n fig(3).

In the graph of the top dark red representation of the below portion of the graph(3-D Sigmoid Curve) .The orange band represents the output as seen in the above graph

As we go along the we will have different colours

> **Here the sigmoid function value lies from 0 to 1.**

> **And the Probablity is also between 0 to1**.

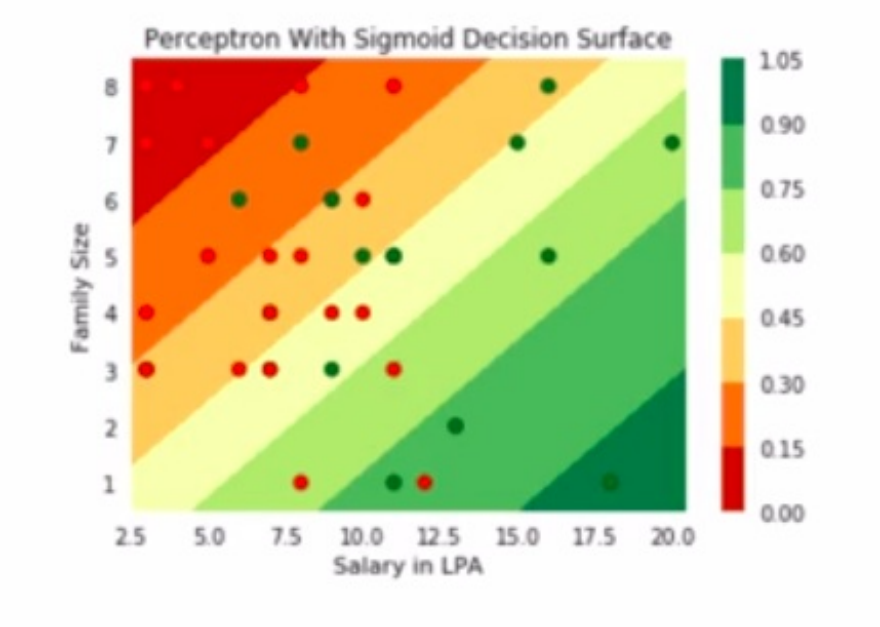
We can interrupt the sigmoid function output as the probability.

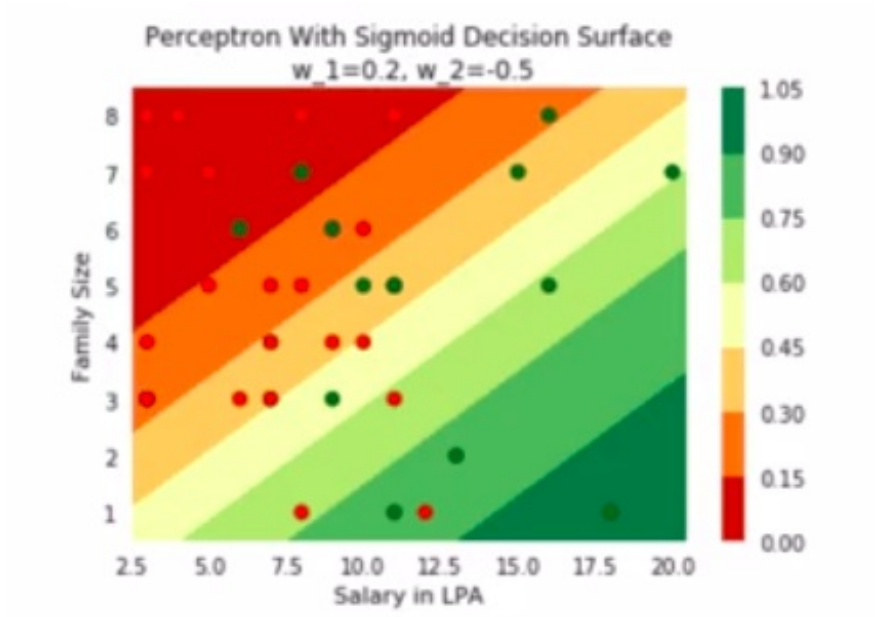
Here in the graph, when it comes to orange then its 0.30 its a 30% chance of buying a car.

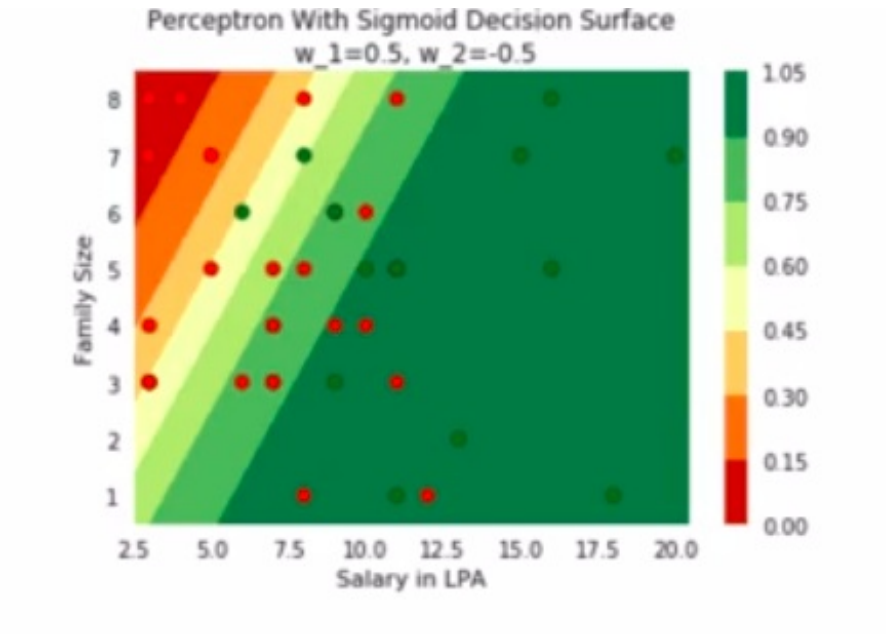
When it comes to dark green it is a 0.90 which is the 90% of buying a car

Like that it varies for different colours like 70%, 30%, 15%, 75%,.....

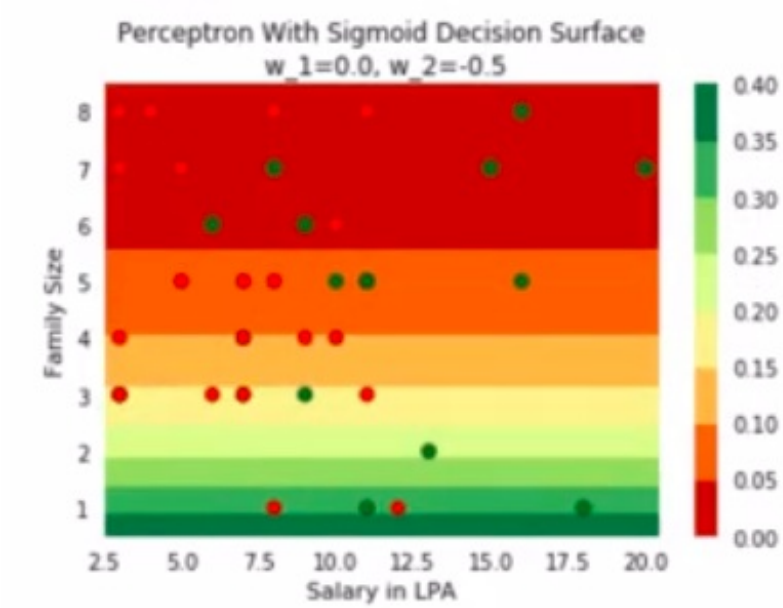
Output is going to change for the sigmoid with change in values.Like different values for w1,w2,b as shown below.

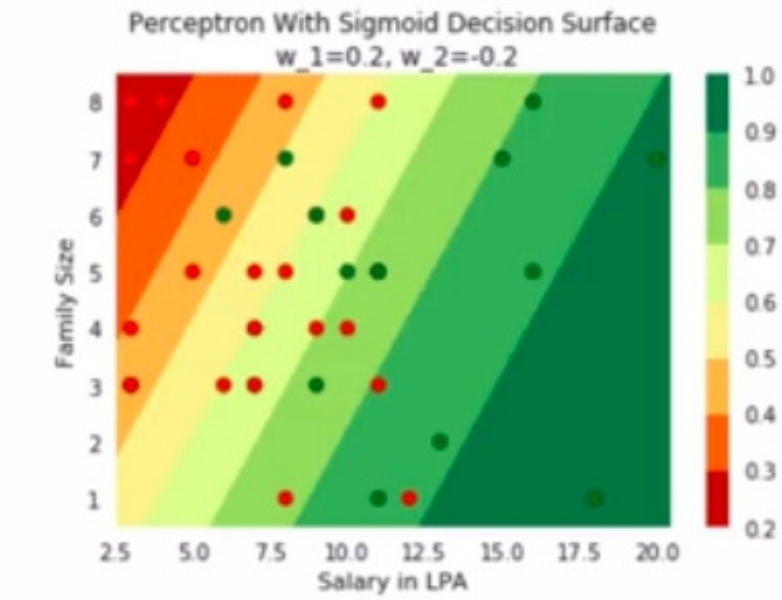




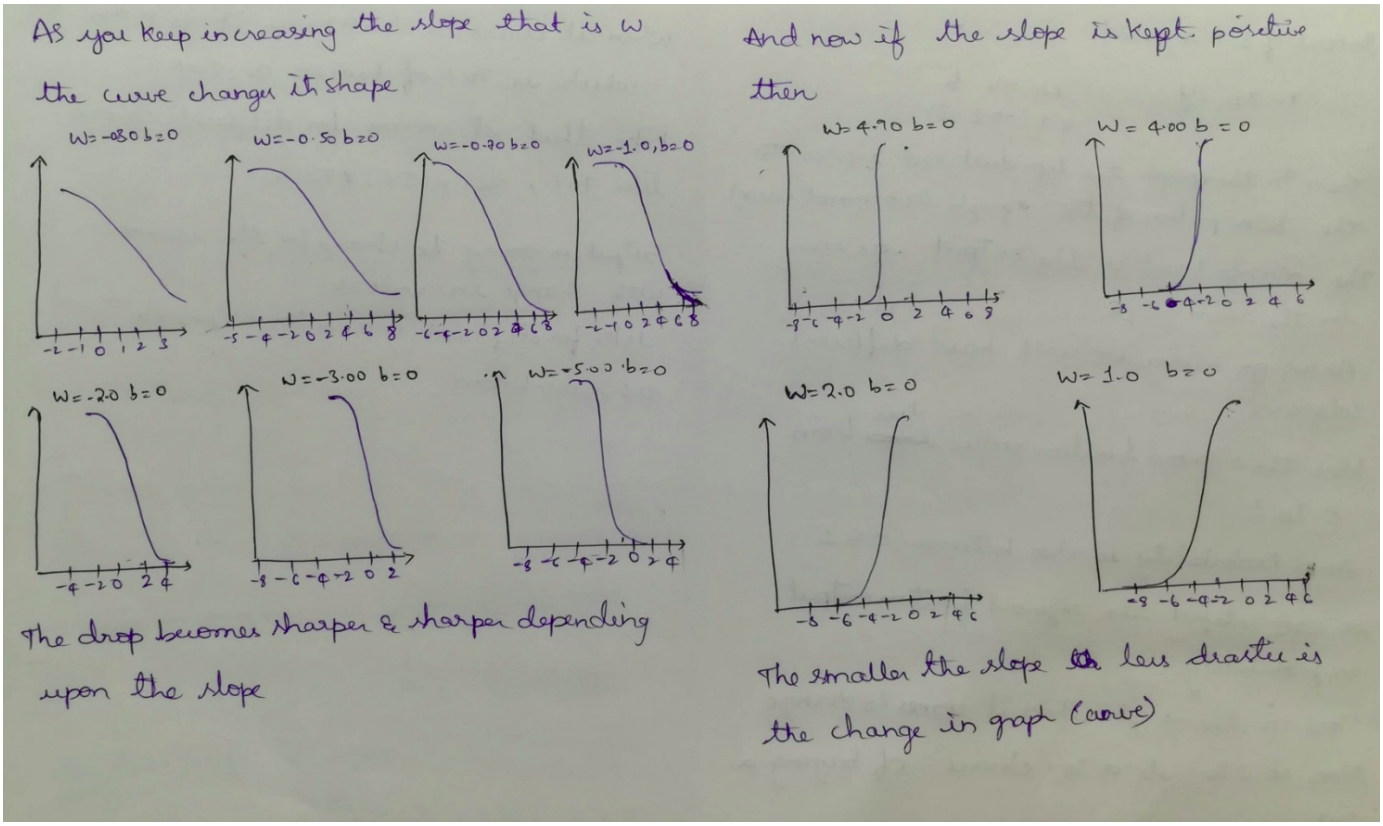


Refer this for the above notes...Like this we can classify data.



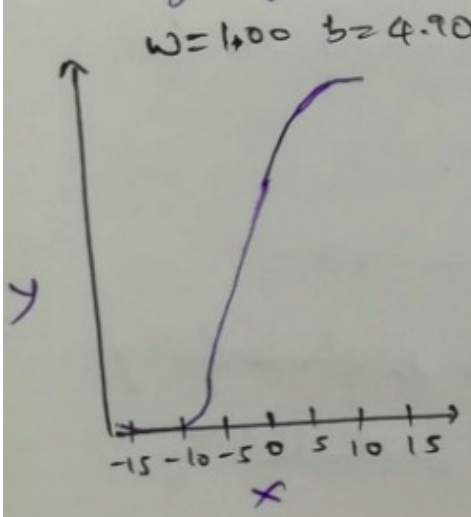


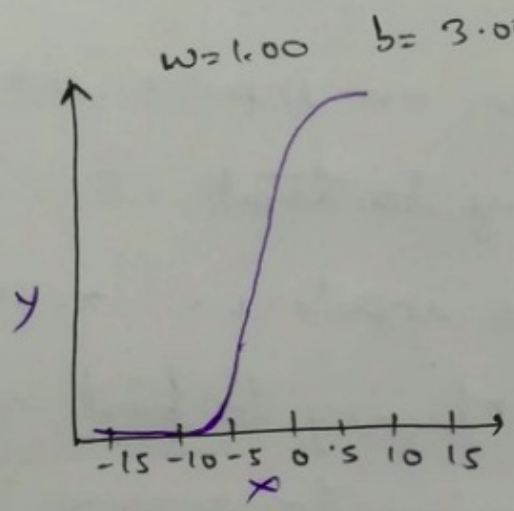
We can classify differently and more accurately compared to Percpetron

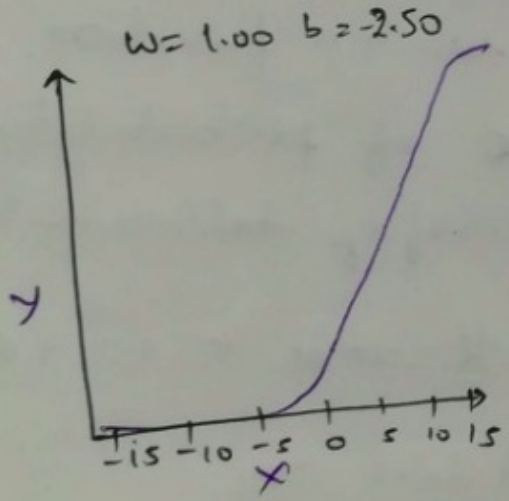


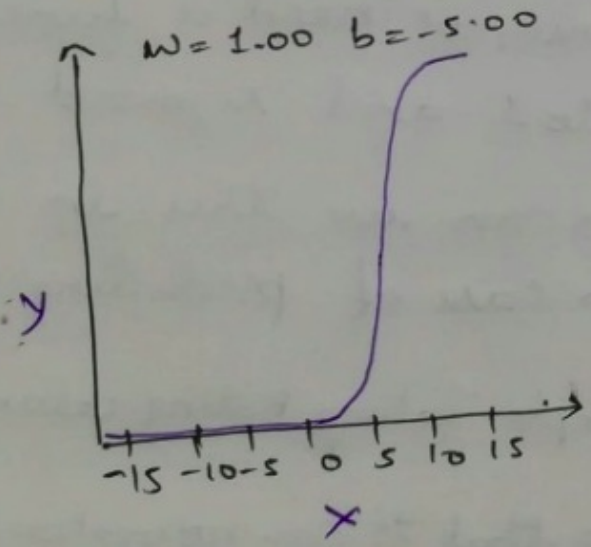
Now coming to change in b values

changing the b value

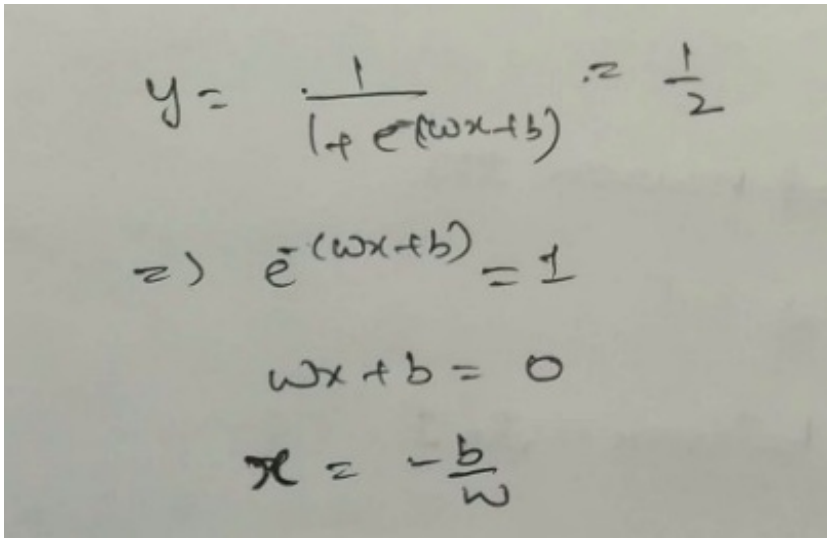








So as u keep decreasing b value the entire funition shifts to right



As you, keep b as more negative and negative the x will be more positive and the whole graph or the curve shifting takes place

This is the proof for the shifts of graph for change in b value.

Data And Task:

Here, for the sigmoid function is y= fhat(x)

here y value is between 0 to 1.

So, we can get any output like 0,7,0.6.0.3,0.2 like this So we take a particular threshold value like 0.5 .

> **If every output >0.5 i will call it as like.**

> **else anything is <0.5 I will call it as Dislike.**

Now, it becomes similar to perceptron but you can ever produce the output flexible without a threshold also.

Regression:

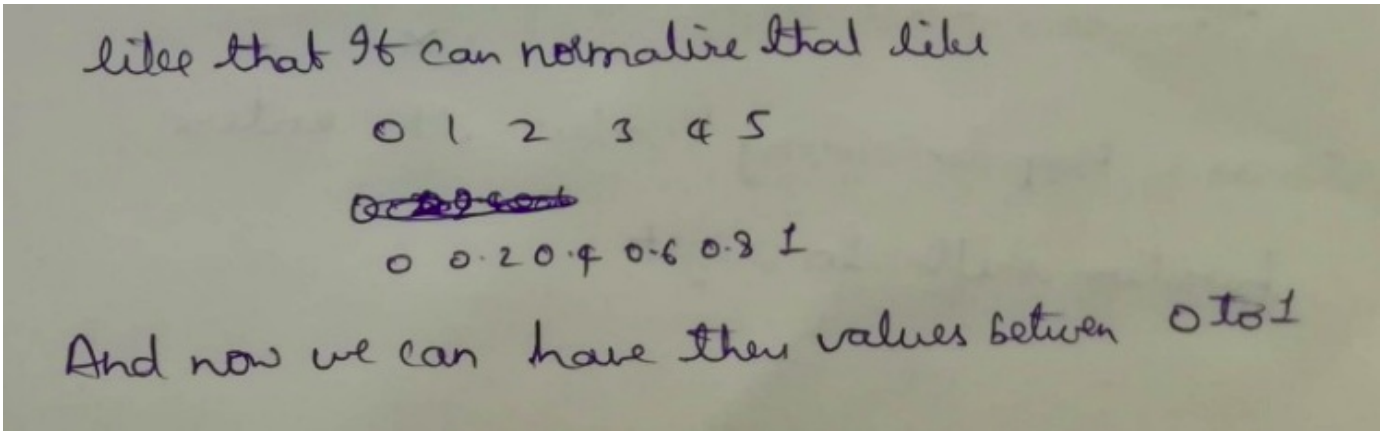
If you are given two variables x and y and given an input x we are going to fit a value y to that.

Given the input i.e., x1,x2,x3,.....,xn.

You are trying to find some probability, you are going regress the value probability.

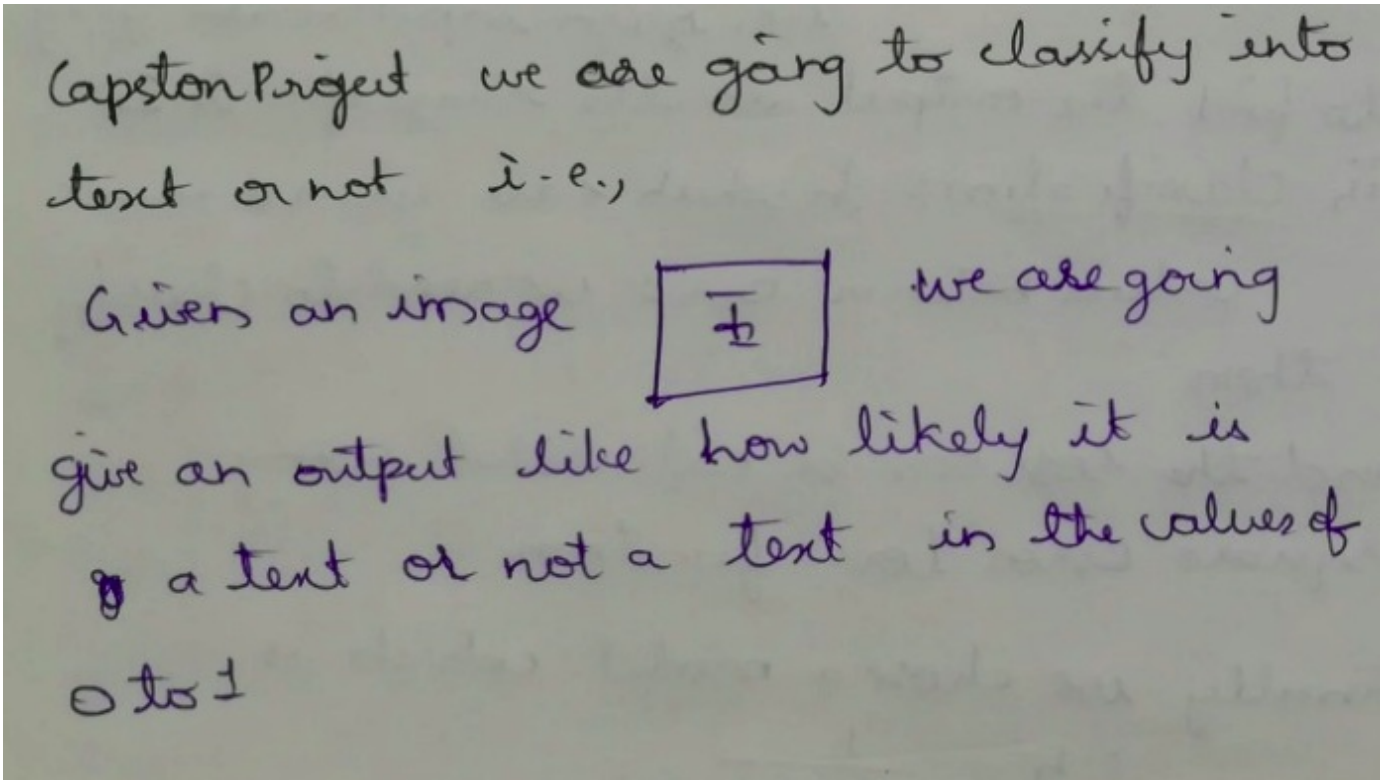
Hence, we need a function which regress from 0 and 1 and sigmoid is such a function.you can use this case, in case of probability predicting values of y between 0 to 1.

> **Example taking Rating column Its can be 0, 1, 2, 3, 4, 5.**



The dataset normalising.

So like you can take a threshold or can give up to ouytput where uer can interpret has +ve and -ve



Capstone Project.

Those are Data and task can be used.

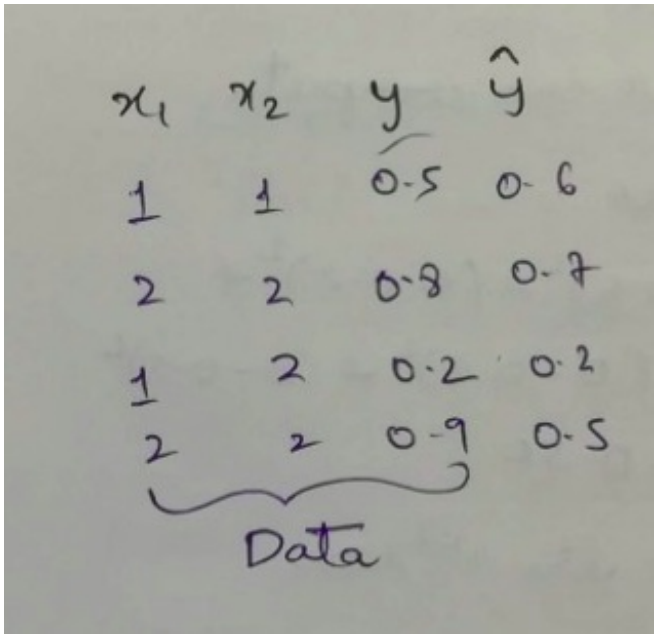
Loss Function:

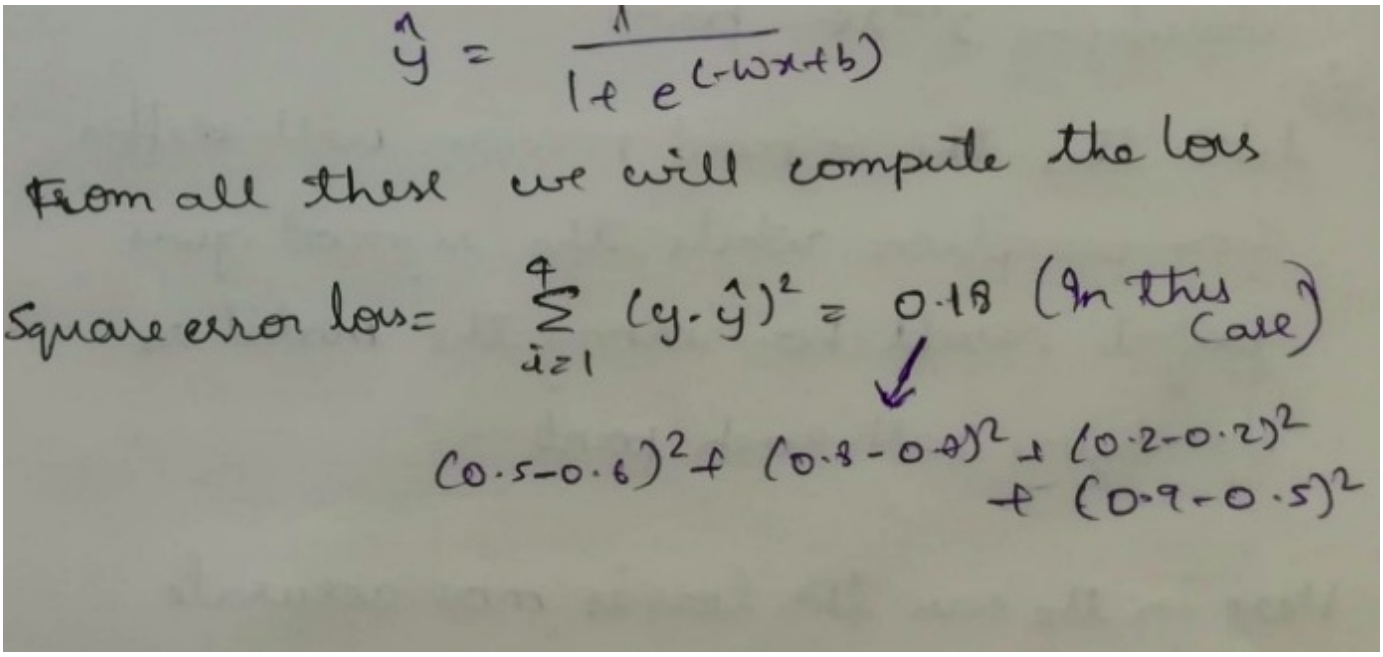
Data set

hetre x1, x2, y are the data which can which we have real any values True output is between 0 to 1

Task again here is regression which is the probability regressing which is between 0 and 1.

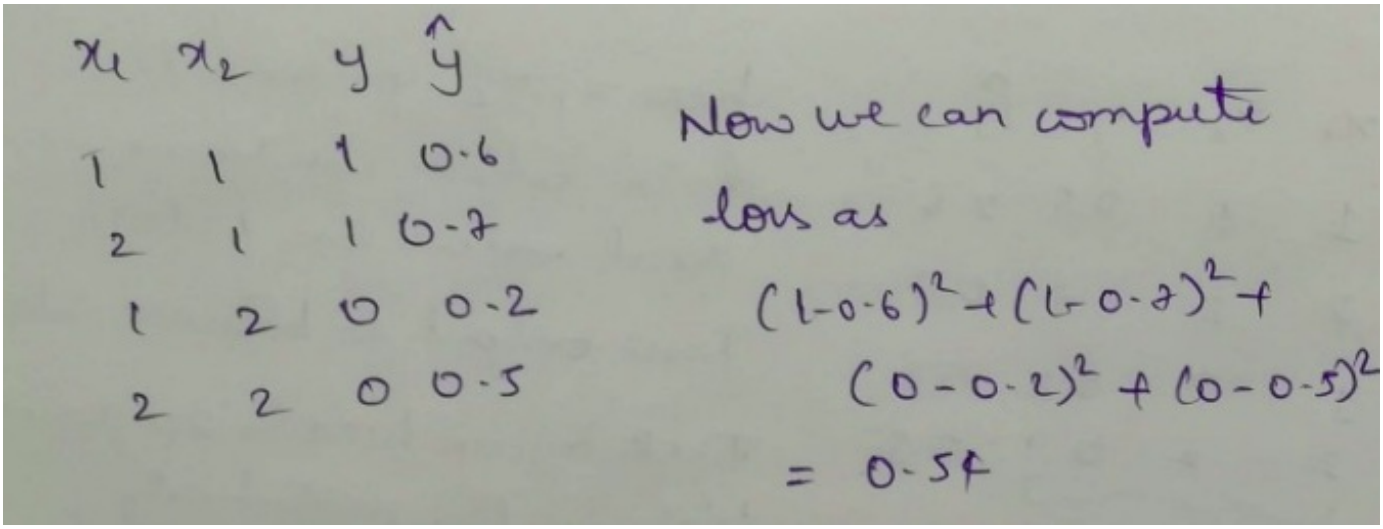
And here the input is related to the output based on logistic function.





Logistic function and calculating Square Error Loss.

If the true value is in the form of binary lets take from that data set and now using mean calculate the loss value.



Calculating loss for the sample data set.

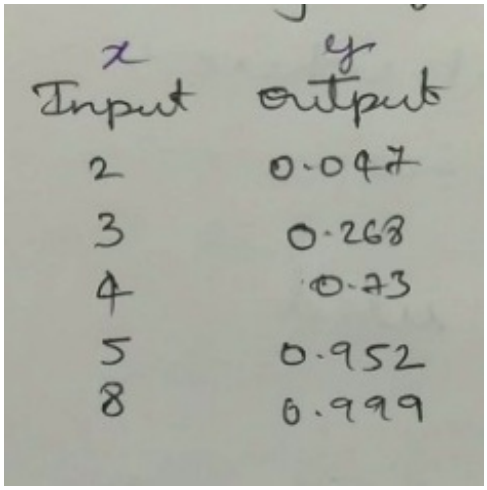
Here, we compare 1st row and the 2nd row then we compare the model is classifying correctly or not 1st row less accurately than classifying 2nd row point

> **Note: Like this the sigmoid neuron will differ from perceptron while the sigmoid gives perfect result how wrong the model is classifying with each point.**

Here in this case the loss is more accurate in sigmoid when compared to the perceptron model.

Learning Algorithm:

Beside is teh data given here we can do two types of Tasks

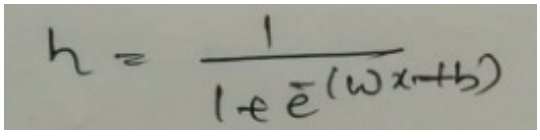


1. **Regression:** In this case, given inputs we are going tp find the outputs in the range of 0 to 1.
2. **Classification:** In which case you are given values from 0 and 1 we need to classify them

And the loss is calculated in these two cases by using the square error loss function.

Finally we chose a model which is sigmoid model

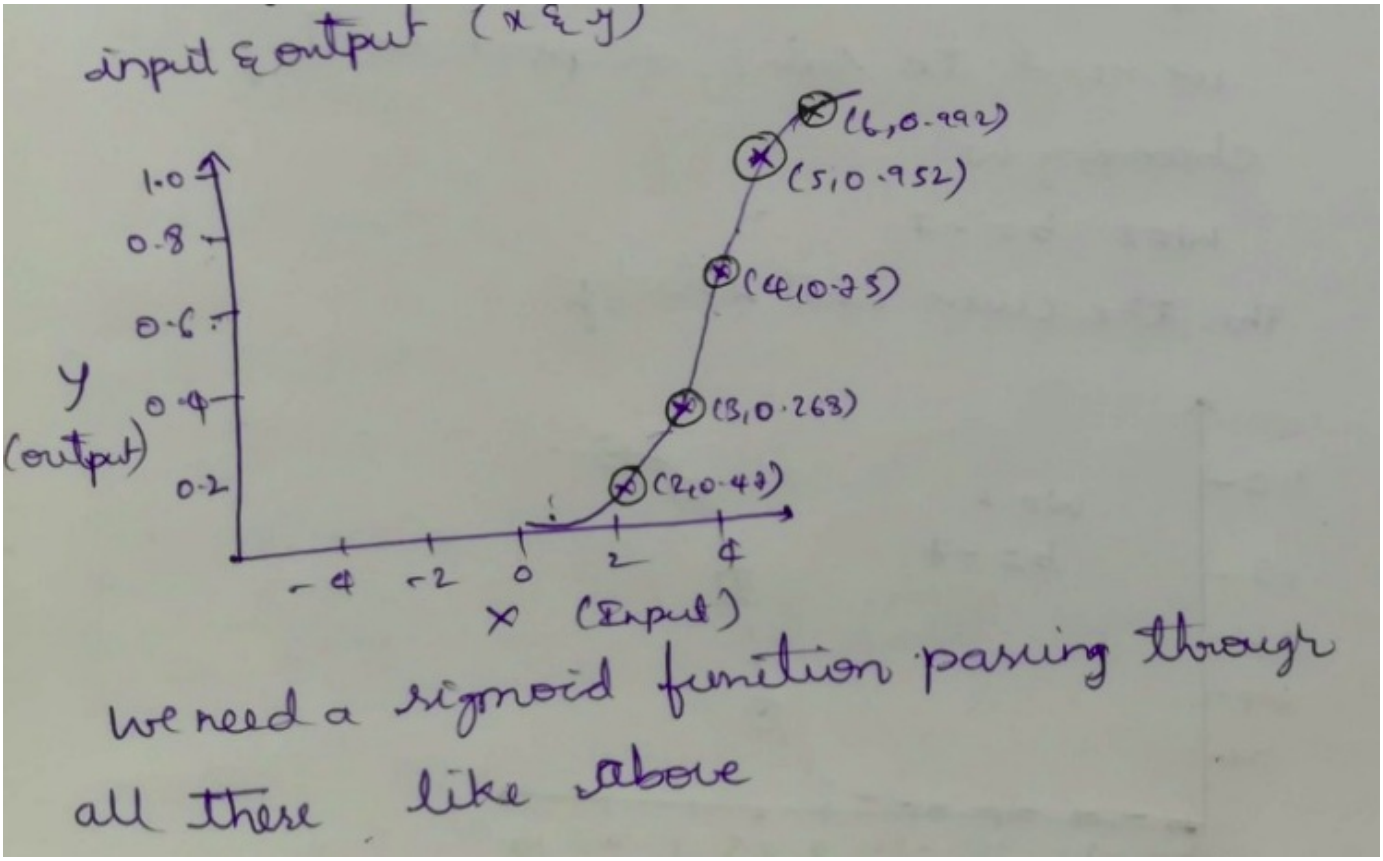
SIGMOID FUNCTION



> **Choosing model implies that for the given x in the table and y given in the table we are going to find the prediction using the beside function i.e., y = fhat(x) parameters w,b.**

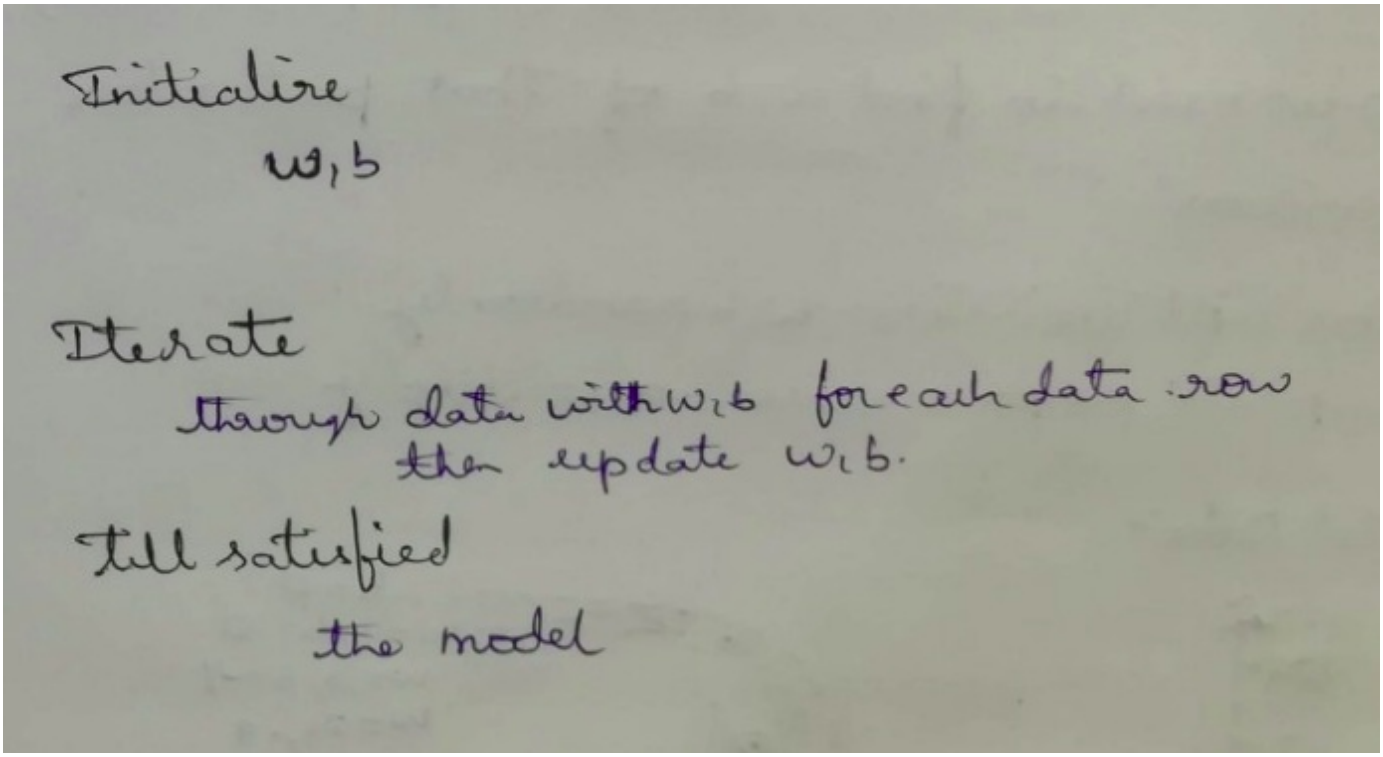
Sigmoid is a **family of functions that means the** by changing w,b values we can get n number of different functions from this family

Actually, we will start plotting points on graph input and output **x and y**.



Plotting the graph.

1. So start with w and b with some values.
2. Go over all the train points.
3. Compute Loss
4. And change the values of w and b accordingly ad we can slowly come to required sigmoid function.



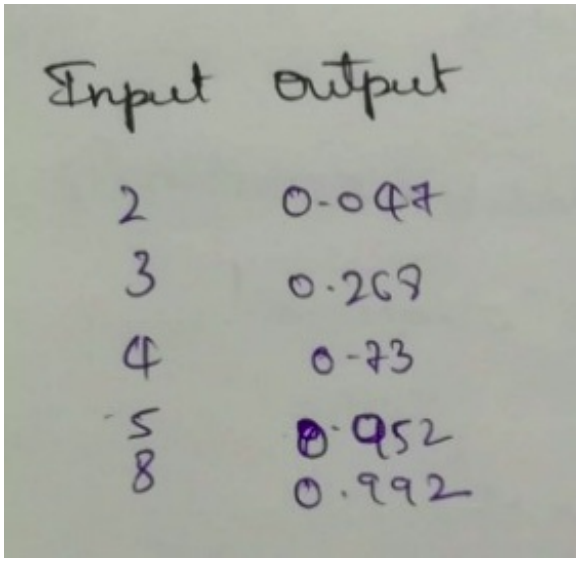
All these above steps are called as simply **Initialise, Iterate, till satisfied phases as show in left side**.

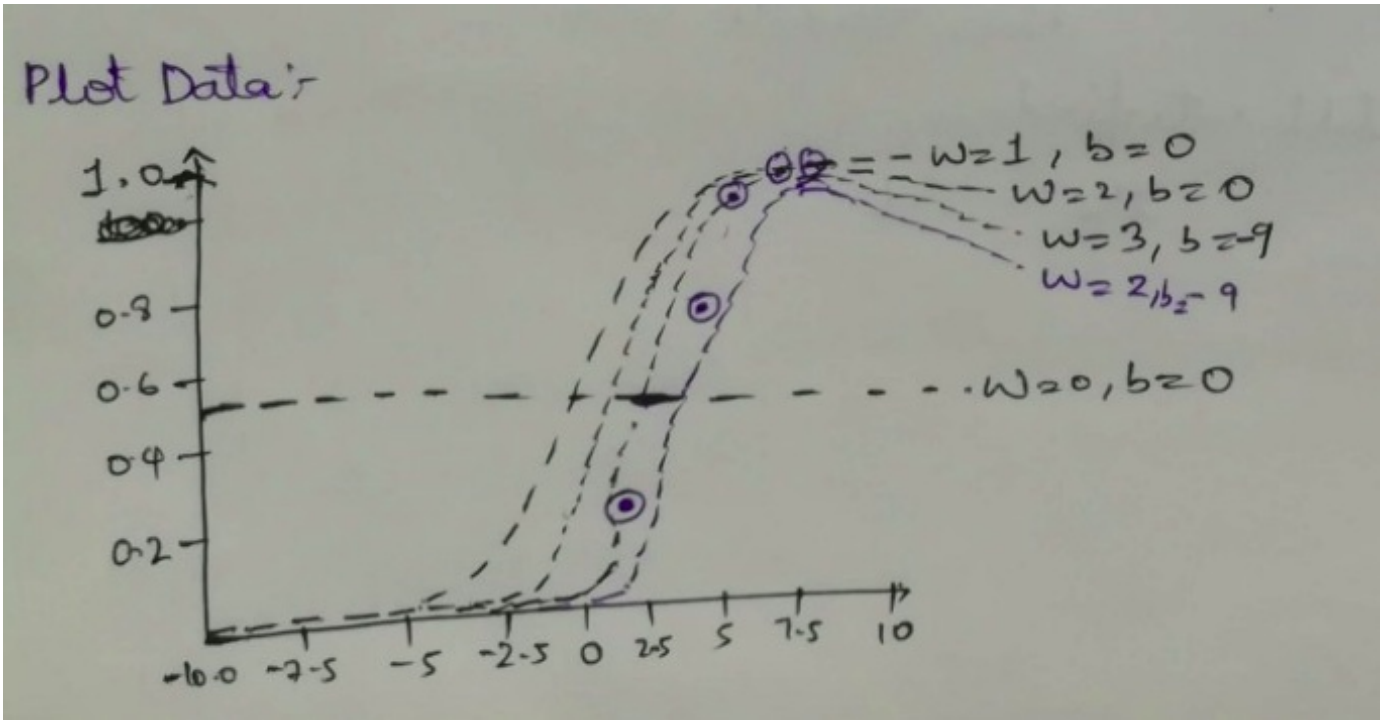
Learning Algorithm Through Guess Work:

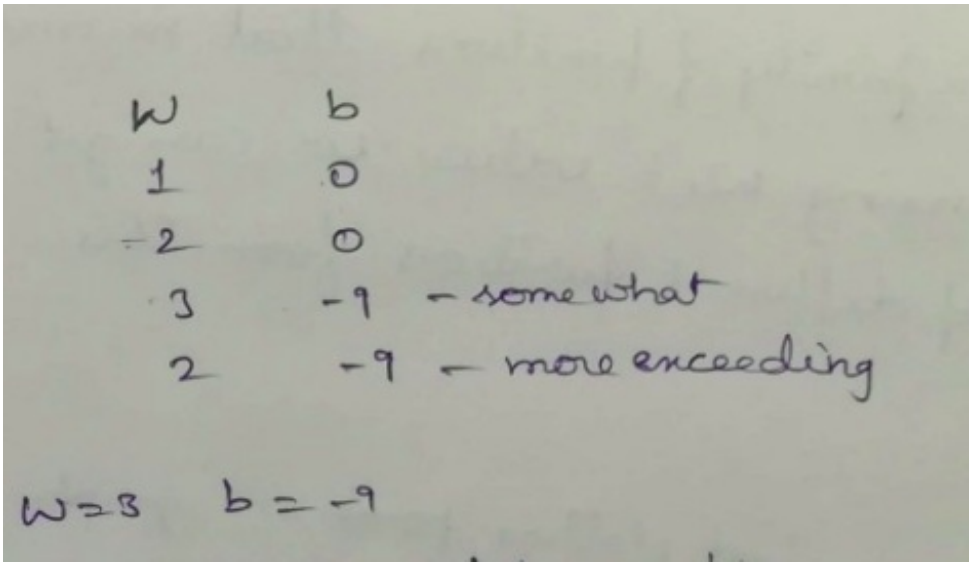
We need a function if we plugin all the inputs like 2,3,4,8,...we need to get the output equals to the beside output(True value).

So, we need to find values of w,b, of the particular function.

> **No we will initialise w,b, randomly update as per required accordingly.**





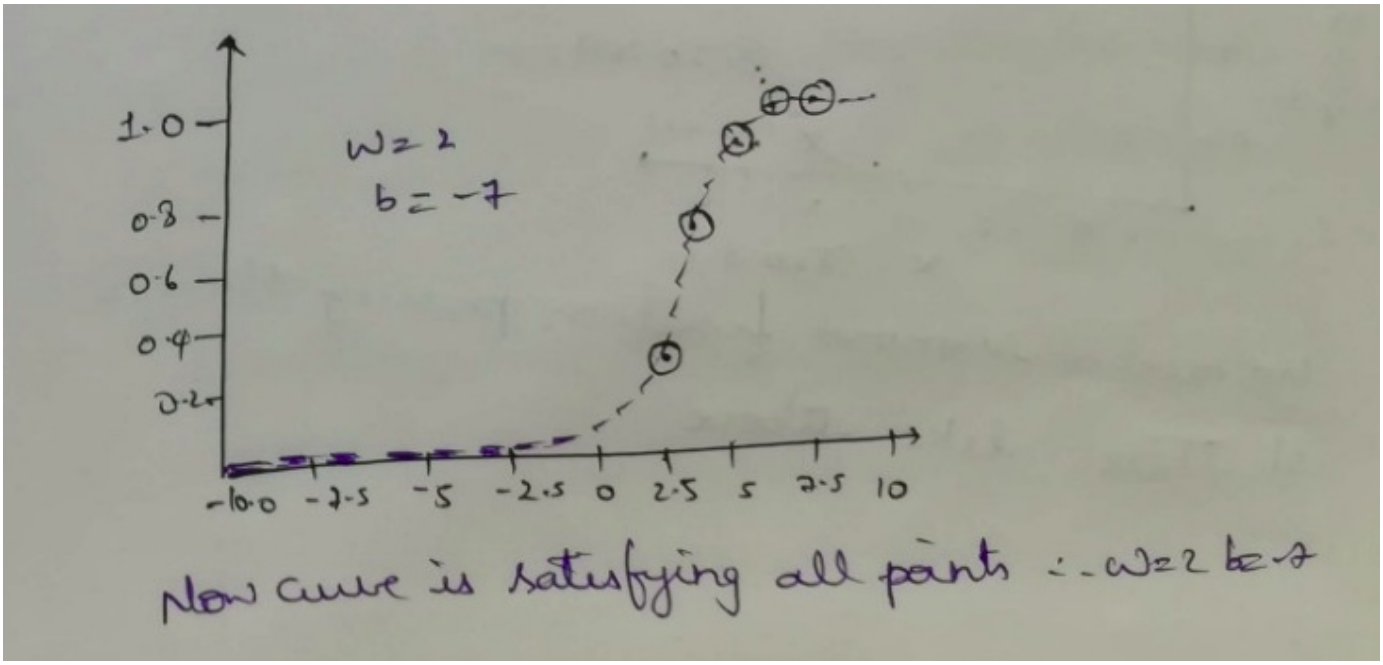


Initialising the w,b randomly and plotting the values

w = 3 and b = -9 is almost satisfying the curve now we need to satisfy all points so try chaging w

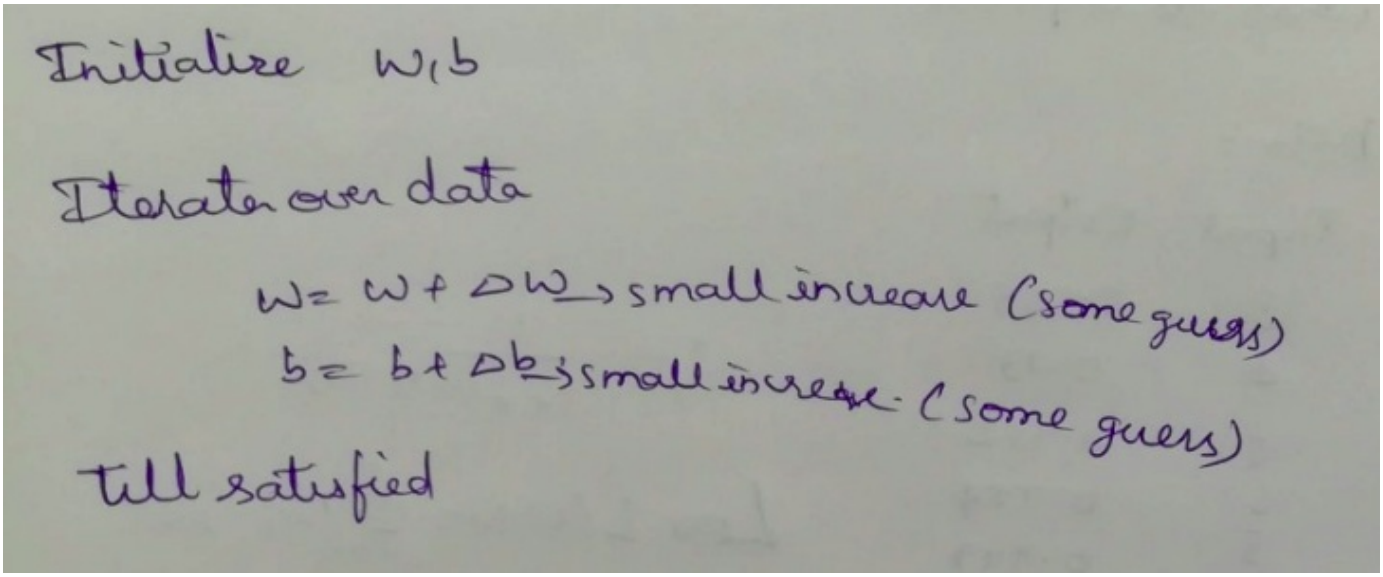
w = 2 and b = -7

We will get the following curve, Now the **curve is satisfying all the points as show in the below picture**.



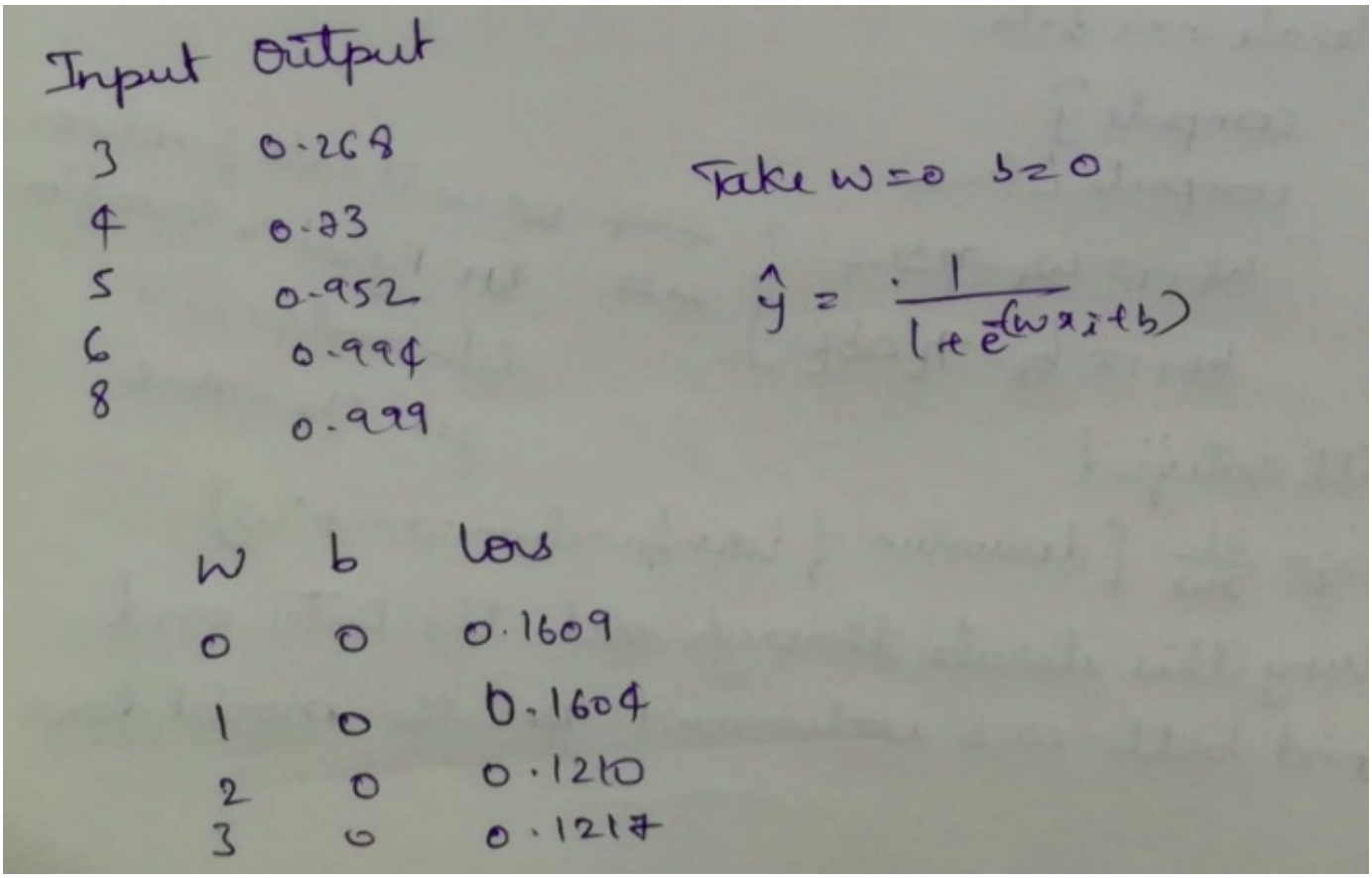
The curve we get for w = 2 and b = -7

What are we doing in guess work:

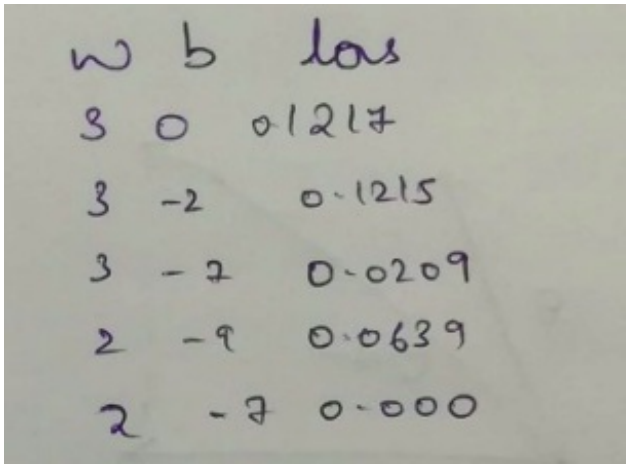


These are steps we follow in the guess work

Loss Function:



Like that as the loss decreases gradually and increasing when we take a false step like that loss value decreases constantly while guessing.



> **Instead of guessing and increasing the values of w and b we need a principle or a procedure to follow because we cant guess correctly every time and we may end up in a infinite loop.**

$\theta = [w,b]$

$\Delta\theta = [\Delta w, \Delta b]$

if we want to change then we need $\theta = \theta + \Delta\theta$

Like the θ new is the new, but a great variation is there between θ and θnew

So we will update $\theta = \theta + \eta\Delta\theta$ **[η is a small value]**

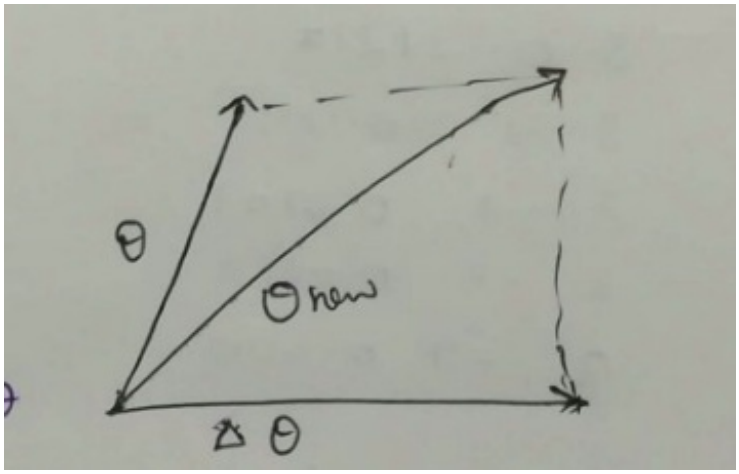
η is any small values just ik ethe learnign rate

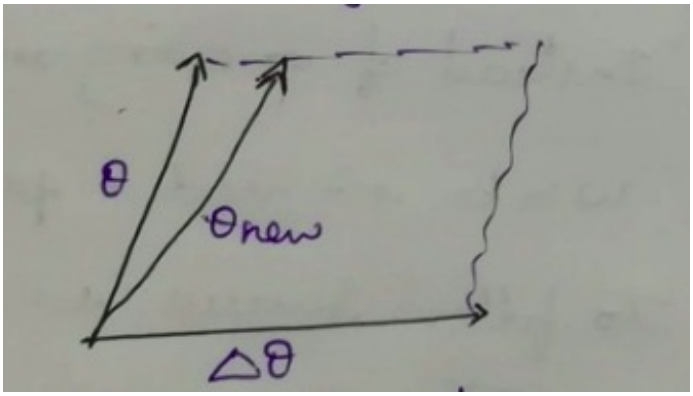
$\theta = \theta + \eta\Delta\theta$

Like this we can change with small value slowly going from one θ value to the another θ.

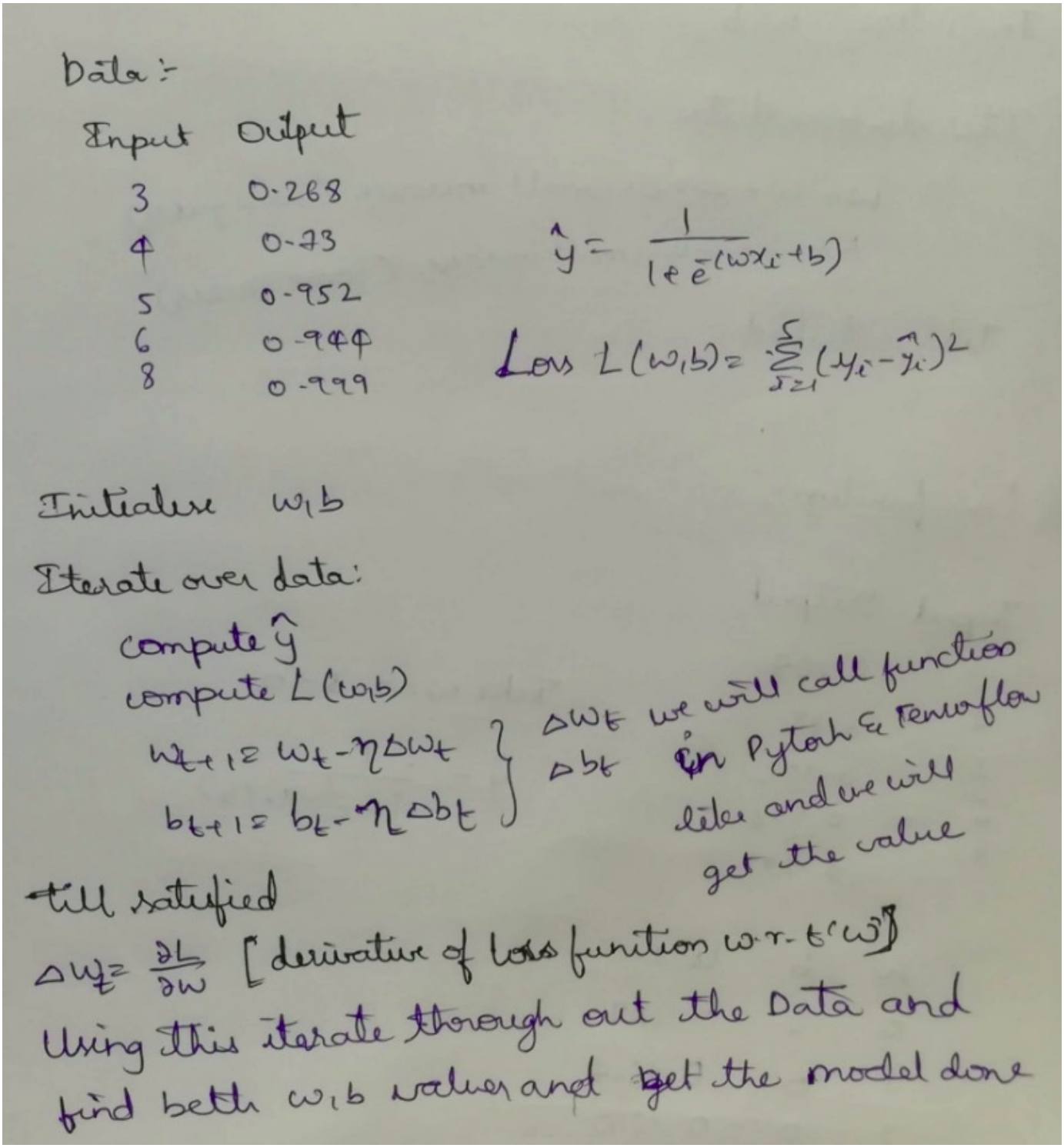
Therefore, from one '**w' to the other** and from **b to the other.**

> **But we always expect Loss(θnew) < Loss(θold).**





How to update? and what to update?



Like this we update he weights.

This all about the Sigmoid Neuron Model ad the working Algorithm.

This is a small try, uploading the notes . I believe in "**Sharing knowledge is that best way of developing skills**".Comments will be appreciated. Even small edits can be suggested.

> **Each Applause will be a great encouragement.**

Do follow my medium for more updates……